DOCUMENT RESUME

ED 171 217 HE 011 301

AUTHOR Owens, Richard E.; Cashin, William E.

TITLE Preparing Examinations: A Practical Guide for IDEA Users.

INSTITUTION Kansas State Univ., Manhattan. Center for Faculty Evaluation and Development in Higher Education.

SPONS AGENCY Kellogg Foundation, Battle Creek, Mich.
PUB DATE 77
NOTE 21p.
AVAILABLE FROM Center for Faculty Evaluation and Development in Higher Education, Kansas State University, 1627 Anderson Ave., Box 3000, Manhattan, KS 66502

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Curriculum; College Faculty; College Students; *Course Objectives; Educational Objectives; Higher Education; Resource Guides; Scoring; Student Testing; *Test Construction; *Test Items; *Test Reliability; *Test Validity

ABSTRACT

The broad topic of preparing examinations for college-level courses is addressed in this guide. General suggestions include the following: devise a test that reinforces the objectives established for the course; decide upon the primary purpose of the test before constructing it; make the test as valid and reliable as possible; and standardize the test. Construction of the following types of items are discussed: recognition items in general, multiple-choice items, true-false items, matching items, recall items, and subjective items. An examination question should be written to accomplish a specific task; it is necessary to consider the characteristics of the course content as well as the level of thinking being tested. Strengths and limitations for each type of item are covered. Suggestions for scoring and an annotated list of further reading materials are included. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED171217

PREPARING EXAMINATIONS

A Practical Guide for IDEA Users

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Center for Faculty Eval. & Dev. HE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Richard E. Owens

William E. Cashin

Center for Faculty Evaluation and Development
in Higher Education
1627 Anderson Avenue, Box 3000
Manhattan, Kansas 66502

HE 011 301

Copyright © Center for Faculty Evaluation and Development in Higher Education, 1977.

The Center is supported in part by a grant from the W. K. Kellogg Foundation of Battle Creek, Michigan.

Permission is hereby granted to reproduce single copies of this Guide for non-profit, instructional use.

Multiple copies may be reproduced by institutions having Memorandums of Agreement with the Center during the period of the Memorandum for non-profit, instructional use within that institution.

Other individuals or institutions wishing to reproduce multiple copies for non-profit, instructional use should obtain written permission from the Center.

Introduction to "Practical Guides"

A major flaw with most student evaluations of teaching, or any evaluation of teaching, is that, having pointed out where there might be room for improvement, they have little to offer about possible means of improvement. These "practical guides" are an attempt to remedy that situation for IDEA users. Typically, the IDEA Report identifies specific teaching methods needing improvement which are related to the instructor's course objectives where the students reported average progress or less. These practical guides will make concrete suggestions about ways to improve teaching. Although the suggestions presented in these guides are written in the context of the IDEA System, most of the suggestions are taken from the general literature on college teaching and so should be of help to any college teacher seeking to improve upon a given teaching technique.

A basic assumption which underlies these guides, and the IDEA System as well, is the belief that there is no SINGLE, CORRECT way to teach. The effectiveness of one's teaching depends upon a multitude of factors including: the objectives of the particular course; the teaching style and other traits of the instructor; and such student characteristics as learning ability, background, motivation to take the course, educational goals, etc. Because of this, these guides will suggest various alternatives and leave to the individual instructor the selection of those which seem more promising.

It is strongly recommended that each instructor employ a variety of methods to obtain feedback concerning the relative success of the alternatives attempted. We particularly recommend that you discuss your attempts to improve your teaching directly with your students. They are, after all, the people who are supposed to be learning.

February, 1977

William E. Cashin
Stephen C. Brock
Richard E. Owens

PREPARING EXAMINATIONS

A Practical Guide for IDEA Users

This guide is concerned with the broad topic of preparing examinations for college-level courses. Three items on the IDEA Survey Form are concerned with examination practices:

6. Gave examinations which stressed unnecessary memorization.

12. Gave examination questions which were unclear.

19. Gave examination questions which were unreasonably detailed (picky).

These items only sample the area of preparing examinations and future research may find additional or more useful items to examine this important area of teaching. Therefore, we have written this guide for college teachers who wish to improve their general examination procedures and not just for instructors who may have received "weak" ratings on one or more of the three IDEA items.

This guide is divided into three parts:

I. GENERAL SUGGESTIONS

- A. Objectives
- B. Purpose
- C. Validity
- D. Reliability
- E. Standardization

II. CONSTRUCTION OF SPECIFIC TYPES OF ITEMS

- A. General Considerations
- B. Recognition Items in General
- C. Multiple-Choice Items
- D. True-False Items
- E. Matching Items
- F. Recall Items
- G. Subjective Items

III. FURTHER READINGS

Instructors whose IDEA Reports suggest improvement related to item 6 should refer to General Suggestions, especially sections A & C. Help with items 12 and 19 can be found in the section on Specific Types of Items.

Preparing examinations is one of two topics where an extensive literature concerned with practical application already exists in higher education (the other topic being stating objectives). For this reason we have omitted concrete examples in writing this guide. Interested readers are referred to Part III Further Readings which lists a number of references containing practical examples.

Perhaps the most frustrating part of the teaching-learning situation for both the teacher and the learner is the classroom test. Students do not seem to know how to prepare for and take tests as effectively as they should; and teachers have generally had little, if any, preparation in the construction and scoring of tests. Yet, this aspect of the teaching-learning process is vital in determining both student progress and the instructor's success in teaching the material. Students may do poorly on an exam because they have not prepared well enough; but if most or all of a class does poorly, it could be the result of a badly constructed test, inadequate instruction, or a combination of all three problems. The test provides both the teacher and the learner with feedback as to how well they have accomplished their tasks.

It is also important to note that tests are vital to student learning not only because they provide feedback but also because students will pattern their study habits according to how they are evaluated. For them, good test scores often become the objective of the course, and they quickly learn what kind of studying results in better scores. If we want students to acquire certain skills in thinking, analyzing, evaluating, etc., our tests must assess those activities.

I. GENERAL SUGGESTIONS

The following suggestions are directed to the more common problems teachers encounter in preparing tests:

A. Make sure the test reinforces and evaluates the OBJECTIVES you have established for the course. If we say our objective is to teach students to think but our tests require only rote memorization, we are neither evaluating thinking ability nor encouraging the students to indulge in it. (See Gronlund, 1970; & Mager, 1962.)

B. Decide upon the primary PURPOSE of the test before constructing it. If the main purpose is instructional, the test should be diagnostic -- designed to help students see their strengths and weaknesses and make decisions about what needs to be done to improve. If the main purpose is mastery, the test should indicate whether the student has reached the minimum level of learning necessary to proceed. When the criterion is not met, the student would normally be expected to study further in the area before going on. If the main purpose is measurement, the test should assess how well the student does in comparison with others who took the test. Since measurement tests judge the relative standing of the student, they are usually designed to spread scores for clearer discrimination into grade categories. Of course, an instructor may choose to work toward more than one purpose; and it is hoped that all tests have some instructional value.

C. Make the test as VALID as possible. Validity refers to how well the test actually measures what it was intended to measure and requires that the test be appropriate for these students and this course as taught this time! Some guidelines to consider are:

1. Attempt to give approximately the same proportional weight to each learning domain as you have in your objectives and your teaching. Three domains of learning have been defined. (Bloom et al, 1956; Krathwohl et al, 1964; Sanders, 1966). Briefly, the Cognitive domain emphasizes the mental processes, ranging from the learning of basic facts to the levels of abstract thinking; the Affective domain comprises interests, attitudes, values, appreciations and individual adjustments; the Psychomotor domain includes skills and physical tasks which require neuro-muscular coordination. For example, if it is important that a student be able to use certain equipment with accuracy, then a "test" should evaluate psychomotor ability as well as the student's knowledge of the necessary manipulations.

2. Balance the weighting of the levels of learning within these domains, making sure that what is asked of the student on the test is similar to what has been expected in class and/or homework assignments. For instance, if an instructor devotes almost all of the class time to the development of formulas but then constructs a test which deals chiefly with application of those formulas to research projects, it is likely that many students will not do well. They will have had little experience with application, and the instructor's teaching in class will not have led them to expect a test of that nature.

3. Test the various course topics in proportion to the emphasis they have received in class. Since tests only sample what students have learned, we often find that they have poor balance; certain topics or chapters are over or under-emphasized on the exam. If there is a deliberate attempt to do this, be sure to advise students of your intentions well in advance.

4. Be cautious about using tests written by others. Textbook tests, tests prepared by outsiders, tests written for several sections of the same course and tests written by a team of instructors have unique advantages, but they can easily be inappropriate for a particular section of a course.

5. Make every effort to assess that which has been accomplished in this course. It is impossible in some areas to avoid testing knowledge accumulated from several courses, but make sure this is not done on all tests. Secondly, be careful not to give tests which are really intelligence tests rather than achievement tests of the material learned in the course. In our attempts to write challenging questions so that students can demonstrate their capabilities, we may require that students use the materials in ways not made explicit in the learning sessions. In such cases the students' intelligence may be a greater factor than how much has been learned in the course. Third, we must be careful about rewarding the "good" test-taker. Part of getting a good score on a test is being able to use effective test-taking techniques, but the premium for this should be small. Such things as answers to previous questions in the stems of later questions will be noted by test-wise students but not by others. It is important to avoid anything which might determine the students' scores other than their mastery of the material.

D. Make your tests as RELIABLE as you can. Reliability refers to the consistency of the test's assessment; if it were possible to erase the students' memories of the test items and they could take the test again, they should obtain similar scores. Anything that increases the "chance factors" in student responses or instructor scoring will lower the test's reliability. Specific suggestions which might help are:

1. Adapt a length that is appropriate for the time actually allowed for the test. Although a class period may be fifty minutes long, the test period often is less. It takes time to distribute materials and establish procedures. Rule-of-thumb guidelines suggest that students can answer about 1 Recall, 1½ Multiple-Choice, 3 True-False or 4 Matching questions per minute.

2. Work to attain a reasonable overall difficulty level as well as a difficulty level appropriate for the type of question used. The overall difficulty of an exam is important because of its impact on students -- they are not as likely to feel encouraged when overall percentage scores are very low, even if the letter grades are satisfactory, i.e., "C" or higher. Also, it is desirable to have some variation in the difficulty level of items. Questions should range from quite easy (to encourage students and relieve some tension at exam time) to very difficult (to challenge and identify the top students). It should be noted that items so easy that 90% of the class answer correctly or so hard that 30% or fewer answer correctly are of little value for measurement. Within this range, each item type has an optimum range -- the variations due to the differing chance factors involved. The majority of items of a given type should

have a difficulty level approximating the figures suggested below if the test is given for measurement purposes: Recall - 40-50%, Multiple-Choice - 55-70%, True-False and Matching - 65-80%. Items used in a mastery test often do not follow these rules since every student is expected to attain certain levels before proceeding.

3. Increase the total number of items and/or tests, if you need to enhance reliability. A modest reliability (.60-.70) on a single 50-question test can be increased to a substantial reliability (.85-.95) by giving 5 or 6 such exams during a semester or quarter. There is, of course, a point of diminishing returns -- one must consider carefully whether time spent for testing might not be more valuably spent in other instructional activity. Also, poor items should never be included for the sake of length.

4. Reliability is negatively influenced by ambiguous questions; tests that are too long or short, too hard or easy; carelessly chosen distractors; inadequate directions; and inconsistent scoring.

E. A good test is STANDARDIZED - every student that takes the test should be given the same opportunity to do well.

1. The directions must be the same for everyone; all should have equal amounts of time, help, equipment, etc.

2. The test should be constructed so that the same performance will receive the same score even if scored by different graders or by the same grader at different times.

3. The test must be easily read -- written both legibly and at the reading level of the students.

4. Special care should be taken to use language similar to that used during instruction -- we sometimes write a bit more formally and at a different level than we speak.

5. Organize and arrange the test lay-out to help the students as much as possible -- the purpose is to test students' knowledge not their eye-sight or puzzle-solving ability.

III. CONSTRUCTION OF SPECIFIC TYPES OF ITEMS

There are three basic categories of test items:

1. Recognition or Objective
2. Recall
3. Subjective

Recognition items include Multiple-Choice, True-False and Matching. In all of these, students are required to select the correct response from the limited number of choices provided. They are "objective" in that the marking process is highly uniform because the key is determined in advance and the same score should be obtained no matter who grades the items.

Recall items involve fill-in-a-blank, short answer, brief one-sentence definitions or explanations, and specific lists of resources or characteristics. Little leeway is allowed in the correct response although some latitude may be allowed for synonyms or slight variations.

Subjective items include Essay items and Oral questions for which no one exact answer is expected. Such items are best used to assess synthesis, creativity, understanding of implications, ability to determine relationships, etc.

Every type of test item has a place in learning evaluation and serves best for certain purposes. Every type of test item also has definite limitations. An examination question should be written to accomplish a specific task; thus, special care is a must when selecting the type of item to use. One has to consider the characteristics of the course content as well as the level of thinking being tested. Meeting these special needs will usually lead to writing a variety of test items.

The following discussion lists strengths and limitations as well as offering construction suggestions -- for each general category (Recognition, Recall, Subjective) as well as for each specific item type.

A. GENERAL CONSIDERATIONS

-Write items as you progress through the semester -- writing a few after each class session is ideal.

-Write items on note cards for ease of handling and flexible use.

-Avoid writing trick items.

-Use a little humor to help relieve student tension, but don't get carried away.

-Keep items as brief as possible or your test may be a speed reading test more than an achievement test.

-Use a variety of item types (e.g., Multiple-Choice and Essay) to help provide for individual differences among students.

-When using a variety of types of items, group the questions according to type so that students don't have to constantly shift response patterns.

-Do not use a series of questions in which answering successfully depends on knowing the correct answer to a previous item.

-Arrange items in order of difficulty; at least, use easier questions for the first few items to avoid unnecessarily discouraging students at the outset.

-Weight the number of points for a question according to the type of item and the amount of learning assessed.

B. RECOGNITION ITEMS IN GENERAL

1. Strengths

-Can be easily and accurately scored by clerk or machine.

-Are versatile -- will adapt to a wide range of content and difficulty levels.

-Are designed to assess the more specific objectives in a course.

-Can provide a comprehensive sample of course content since the response mode is simple and requires little student time.

2. Limitations

-May not allow students to demonstrate a range of knowledge since the response is limited to the options provided.

-Encourage guessing since one of the options is correct.

-Are difficult to write so that all students will have the same interpretation.

-Are not easily used to assess affective learning.

-Are so easily constructed to assess basic factual knowledge that instructors often fail to test higher levels of thinking.

3. Construction Suggestions

-Avoid patterns in the response key. It may make scoring easier, but sharp students will "break the code."

-Avoid over-using any one response as the correct answer -- using far more true than false or more "B" choices regularly gives the student a clue as to which response to bet on in a pinch. For example, McKeachie (1969, p. 134) suggests to students that instructors tend to use "B" options for the correct answer more often than other options.

-Strive to include items which test the higher levels of thinking and learning.

-Avoid such qualifying words as always, never, etc.

-Use the positive statement whenever possible.

-When necessary to use a negative statement, underline the negative word.

-Be careful not to give answers or clues to answers in preceding or subsequent items.

-Use a simple response method -- separate answer sheets, a column of response blanks set to the side of the items, optical mark reader cards or optical scan sheets.

C. MULTIPLE-CHOICE ITEMS

1. Strengths

-Are considered to be the most versatile of all item types.

-Can be used to test level of understanding and ability to apply learning.

-Can be written to assess ability to integrate information from several sources when formulating an answer.

-Are very useful for obtaining diagnostic feedback if the incorrect alternatives are constructed to cover common errors.

-Provide an excellent basis for post-test discussion, especially if discussion includes why the incorrect responses were wrong as well as why the correct responses were right.

2. Limitations

-Are difficult and time-consuming to construct well.

-May appear too discriminating (picky) to students, especially when the alternatives are well constructed.

-Are open to misinterpretation by students who read more into questions than is there.

3. Construction Suggestions

-The stem should clearly present the problem or question.

-It is generally best to use direct questions rather than incomplete statements for the stem.

-Include as much of the item as possible in the stem so that the alternatives can be kept short.

-In testing for definitions, it is best to use the term in the stem rather than as an option.

-The alternatives should be listed on separate lines so that all options can be clearly distinguished. Never include the options as part of a narrative along with the stem.

-All alternatives should be homogeneous in content.

-All options should be plausible responses to the stem. Do not write poor alternatives just for the sake of having more options.

-All choices should be grammatically consistent.

-Keep alternatives for an item approximately the same length. Especially avoid the common error of making the correct response consistently longer.

-When composing incorrect alternatives, keep in mind the misconceptions students have indicated in class or errors commonly made by students in the past.

-Recognize that using "all of the above" and "none of the above" as alternatives will change the item to a series of True-False questions.

-Use capital letters (A,B,C,D,E) as response signs; they are not so easily confused when written as are the small letters (especially "a and d" and "c and e").

-If all items do not have the same number of alternatives, group those of equal numbers so that students do not have to constantly adjust for the pattern.

D. TRUE-FALSE ITEMS

1. Strengths

-Are relatively easy to prepare since each item comes rather directly from the content.

-Offer the opportunity to write questions which cover more content than most other item types since students can respond to so many in the time allowed.

2. Limitations

-May not give a true estimate of the students' knowledge since 50% could be gotten correct simply by chance.

-Are very poor for diagnosing student strengths and weaknesses.

-Are generally considered to be "tricky" by students.

-Tend to be either extremely easy or extremely difficult.

3. Construction Suggestions

-Use a relatively large number of items (75 or more when the entire test is T-F) since a student could correctly guess half by chance.

-Do not take statements verbatim from the text.

-Avoid extremely long and complicated statements, or you may be testing reading ability rather than knowledge of the material.

-Avoid negatives, especially double negatives.

-Avoid ambiguous and trick items.

-The diagnostic value of the items can be increased by having students revise all false statements to read as true ones.

-Use statements which are entirely true or entirely false - otherwise the well-informed student is often penalized.

-Avoid tip-off words: all, always, never, every, none and only usually indicate a false statement whereas generally, sometimes, usually, maybe, often are often used in true statements.

-Avoid ambiguous terms: several, seldom, frequently, etc.; use phrases like "50% of the time."

E. MATCHING ITEMS

1. Strengths

-Are especially suitable for who, what, when and where questions.

-Are generally quite brief and uninvolved.

-Permit more efficient use of space when there are a number of similar types of information to be tested.

2. Limitations

-Seldom can be used to measure beyond recognition of basic factual knowledge.

-Are usually poor for diagnosing student strengths and weaknesses.

-Are appropriate in a limited number of situations.

3. Construction Suggestions

-Include only homogeneous material in a set of matching items (i.e., do not use dates and places in the same set).

-To reduce the amount of reading, use the more involved expressions in the stems and keep the responses short and simple.

-Arrange the items in the response column in some logical order -- alphabetical, numerical, chronological, etc. -- so that students do not waste time searching.

-Always place an entire set of items (stems and responses) on the same page.

-Use no more than about fifteen items in one matching set.

-If responses are to be used only once, provide more responses than stems to prevent students from using the process of elimination rather than concentrating on the answering of each question.

-It is permissable to use a response as the correct answer for more than one stem but never multiple responses for one stem.

-Number each stem for ease in later discussions.

-Use capital letters (A,B,C,D,E) for the response signs.

F. RECALL ITEMS

1. Strengths

-Are very easy to prepare.

-Use vocabulary or phrasing exactly as presented in instruction or readings.

-Are good for strictly factual information where a specific word or expression is important to know.

-Reduce guessing since they require a definite response rather than simple recognition of the correct answer.

2. Limitations

-Reduce objectivity in scoring unless only one word or expression is correct.

-Tend to test only rote, repetitive responses -- often using the words of the text.

-May discourage the student because the task is so precise and often difficult.

-Make scoring difficult and time-consuming -- often must be done by the test writer.

-Encourage a fragmented study style since memorization of bits and pieces will result in higher scores.

-Have little evidence to recommend them over other item types unless the need for specific recall is essential.

3. Construction Suggestions

a. General

-It is best to use original questions rather than take sentences directly from the text.

-Make the statements clear and concise, leaving no question about what is to be done.

-Use vocabulary and phrasing as found in the text or as presented in class.

-Give explicit directions as to what is expected of the student and what variations will be accepted in the answers.

-It is recommended that only partial credit ever be deducted for spelling errors, omitted units on math problems, etc.

-Much more credit should be given for Recall items than for T-F or Matching.

b. Fill-in-the-blank Items

-Word the statement to encourage a specific response.

-Never use a long quote and leave a blank every few words.

-The blank should be a one-word response -- or at least a phrase of closely related words.

-Scoring can be facilitated by having students write their answers in a column to the left of the items.

-Leave out only important terms or expressions for students to fill in. Do not leave out insignificant words such as "a" or "the".

-Use "a/an" preceding blanks in order to avoid giving clues to the correct answer.

c. Short Answer, Definitions and Lists

-Specify the source of the item (lecture, text, etc.) so that students will have the criteria to judge their own responses.

-Indicate whether you require specific wording, paraphrasing or illustrations, etc.

-It is usually best to ask for a specific number in a list and base the credit given on that number. Asking for "5 of 6" gives the student a little margin.

-Leave enough space for the student to respond right after the question -- both to help the student keep the question in mind and to give some guidance as to the length of response expected.

G. SUBJECTIVE ITEMS

1. Strengths

a. General

-Encourage students to strive toward understanding a concept as an integrated whole.

-Permit students to demonstrate achievement of higher level objectives -- such as analyzing given conditions, contrasting and comparing, etc.

-Allow expression of both breadth and depth of learning.

-Encourage originality, creativity and divergent thinking.

-Offer students the opportunity to use their own judgments, writing styles and vocabularies.

-Can be written to test abilities to reason, integrate, think critically and solve problems.

-Are less time-consuming to prepare than any other item type.

b. Oral items have these additional strengths

-Allow students to modify responses according to the examiner's reactions.

-Allow for further probing when an answer is not as explicit as desired by the examiner.

-Eliminate the hazard of judging an answer upon the basis of the writing skill exhibited.

-Are potentially useful with handicapped students.

2. Limitations

a. General

-Permit only a limited sampling of learning due to the time required for students to respond.

-Are not appropriate for every subject matter.

-Are not appropriate for assessing knowledge of basic facts.

-Provide students more opportunity for bluffing, rambling and "snowing".

-Are pitfalls for students who tend to go off on tangents or misunderstand the main point of the question.

-Are very difficult and time-consuming to score.

-Are subject to biased and unreliable scoring.

-Permit ratings influenced more by organizational style, grammar and neatness than by content of response.

b. Oral items have these additional limitations

-Are especially trying for students who have poor speaking skills or are anxious in verbal situations.

-Allow the examiner to be a positive or negative influence.

-Do not allow student anonymity for scoring.

-Present complicated and time-consuming scheduling problems.

-May give an advantage to later students if only one question is used and information is "leaked"; yet, multiple questions of equal difficulty are tough to produce.

3. Construction Suggestions

a. General

-Use novel problems or material whenever possible, but make sure they relate to class learning.

-Be sure the questions are comprehensive rather than dealing with small units of content.

-Be careful to present directions so that the student is perfectly clear about what needs to be done.

-Instruct students as to how to respond -- plan ahead, outline before final response, distribute time cautiously, etc.

-Allow appropriate time for students to do justice to the questions.

-Provide guidelines concerning the amount of time to spend on each question as well as the length of response.

-Specify the proportional value of the questions in comparison to the total grade.

-If you prefer a particular response style (shorter responses, use of phrases rather than sentences, etc.), tell this to the students before the test.

-When dealing with controversial topics, make it clear that the task is to logically present a position rather than take a particular stand.

b. Specific suggestions for oral items

-It is best to develop a pool of items of about equal difficulty and comprehensiveness.

-Have more than one person answer each question.

-Use a random selection process when choosing the specific question for a student.

-Be careful not to argue with the student.

-Avoid such behavior as raising the eyebrows which might signal to the student that he's on the wrong track and should find another answer.

4. Scoring Suggestions

a. General

-Prepare, in advance, some acceptable responses -- at least, list the key points you expect to be made.

b. Essay Items

-Grade the papers as anonymously as possible -- have students put their names only on the back of the last page.

-Grade one question on all papers before proceeding to the next item.

-Before marking any paper, read through several responses to a question in order to get a general feel for the norm.

-After grading one question, shuffle the papers so that no paper is consistently scored first or last.

-Be liberal with comments, emphasizing strengths as well as weaknesses, along with suggestions for improvement.

c. Oral Items

-Use every possible means to avoid favoritism in your evaluation.

-After all have been examined, be sure to give the students feedback in writing -- comment on strengths and weaknesses, make suggestions.

FURTHER READING

As was stated earlier there is an extensive literature in higher education which provides concrete, practical help for the college teacher who wishes to improve his or her examination procedures. Many excellent references are NOT included on the list which follows, but we have found those listed to be helpful. All of the references below contain PRACTICAL EXAMPLES.

Bloom, B.S., et al. (Ed.) Taxonomy of educational objectives: cognitive domain. New York: David McKay, 1956.

A very detailed treatment of the work of the committee which originally developed these cognitive objectives. The taxonomy was primarily developed to classify test questions but has been applied as a useful frame of reference for considering educational goals in general.

Educational Testing Service. Making the classroom test: a guide for teachers. Evaluation and Advising Service Series, No. 4. Princeton, NJ: Educational Testing Service, 1961.

A pamphlet which discusses and illustrates the basic rules of test making, compares essay and objective test questions, and describes and illustrates good test items.

Educational Testing Service. Multiple-choice questions: a close look. Princeton, NJ: Educational Testing Service, 1963.

A pamphlet specifically devoted to writing multiple-choice items which require thought, insight and understanding. It is primarily a set of illustrations of thought-provoking multiple-choice examples.

Engelhart, M.D. Improving classroom testing. Washington, D.C.: National Education Association, 1964.

A pamphlet which emphasizes the use of objectives in preparing tests, discusses the use and preparation of various item types, presents some information on grading and interpreting results. The examples are mainly at the high school level.

Gorow, F.F. Better classroom testing. San Francisco: Chandler, 1966.

A brief book designed specifically to help teachers design classroom tests in line with the course objectives, construct test items that measure learning beyond the knowledge level, improve analysis of test results, and become more expert in test construction. The examples are mainly from high school courses.

Green, J.A. Teacher-made tests. New York: Harper and Row, 1963.

A book which stresses practical applications rather than theory with sections on construction and use of test items, scoring and grading, interpreting results. Directed to primary, secondary and college-level teachers.

Gronlund, N.E. Measurement and evaluation in teaching. (3rd ed.) New York: MacMillan, 1976.

A standard textbook. Part II. Constructing Classroom Tests has six chapters related to classroom examinations.

Gronlund, N.E. Preparing criterion-referenced tests for classroom instruction. New York: Macmillan, 1973.

A 52-page description of criterion-referenced testing and detailed discussion of steps involved in constructing such tests. Most of Gronlund's suggestions have applicability to other classroom tests as well.

Gronlund, N.E. Stating behavioral objectives for classroom instruction. New York: Macmillan, 1970.

A 58-page companion booklet to "Preparing criterion-referenced tests" and necessary prior reading because without clearcut objectives, testing has little meaning.

Krathwohl, D.R., et al. (Ed.) Taxonomy of educational objectives: affective domain. New York: David McKay: 1964.

A continuation of the work of Bloom et al (1956) dealing with affective (attitudinal) objectives. Describes the affective levels and gives examples.

McKeachie, W.J. Teaching tips: a guidebook for the beginning college teacher. (6th ed.) Lexington, MA: D.C. Heath, 1969.

Still probably the best, single, practical book on college teaching appropriate for all college teachers, the subtitle notwithstanding. Contains chapters on almost every issue of concern to college teachers including one on examinations.

6

Mager, R.F. Preparing instructional objectives. Palo Alto, CA: Fearon Publishers, 1962.

This is probably the classic book on behavioral objectives. Its step-by-step, leading-by-the-hand approach offends some college teachers, but the book is well worth the hour or two it takes to work through.

Milton, O., & Edgerly, J.W. The testing and grading of students. New Rochelle, NY: Change Magazine, 1976.

An up-to-date (1976) discussion of the broad issues of testing and grading in college. Contains a chapter on constructing tests.

Popham, W.J. Criterion-referenced measurement: an introduction. Englewood Cliffs, NJ: Educational Technology Publications, 1971.

A book of six papers on criterion-referenced measurement, ranging from basic concepts to reports of statistical studies on the topic.

Sanders, N.M. Classroom questions: what kinds? New York: Harper & Row, 1965.

This book, which uses examples from high school social studies, is concerned with levels of questions (according to Bloom's Taxonomy). Sanders' book actually provides a more understandable and usable approach than the Taxonomy itself.

Thorndike, R.L. (Ed.) Educational measurement. (2nd ed.) Washington: American Council on Education, 1971.

Technical reference on most aspects of educational measurement. Chapters on "Planning the Objective Test", "Writing the Test Item", and "Essay Examinations" are recommended.

Thorndike, R.L., & Hagan, E. Measurement and evaluation in psychology and education. (3rd ed.) New York: John Wiley & Sons, 1969.

A standard textbook. Chapters on "Planning a Test" and "Preparing Test Exercises", cover the basics in constructing classroom tests.

Wood, D. Test construction. Columbus, OH: Charles E. Merrill, 1961.

A comprehensive guide for teachers on classroom testing. It includes sections on the principles involved in good test writing; descriptions, uses, and construction suggestions for various item types; treatment and analysis of test data. Directed to teachers at all levels.